SIGNIFICANCE OF NORMS AND COMPLETENESS IN VARIATIONAL BASED METHODS

Joel A. Storch

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA

Abstract

By means of a simple structural problem, we bring into focus an important requirement often overlooked in practice on the basis functions used in Rayleigh-Ritz-Galerkin type methods. The problem of the static deformation of a uniformly loaded beam is solved variationally by expanding the beam displacement in a Fourier Cosine series. The potential energy functional is rendered stationary subject to the geometric boundary conditions. It is demonstrated that the variational approach does not converge to the true solution. The object of the paper is to resolve this paradox, and in so doing, indicate the practical implications of norms and completeness in an appropriate inner product space.

Introduction

Virtually all flexible multibody codes in use today are based upon some variational principle of mechanics. The most common of these being Hamilton's Principle and its discretized version-Lagrange's Equations. Regardless of the particular label attached to the technique (e.g., assumed modes method, Ritz-Galerkin), the problem reduces to rendering stationary a certain definite integral with respect to a sufficiently regular family of functions subject to certain "geometric" boundary conditions. In practice the basis functions are generated using a general purpose Finite Element program and may be subject to further manipulation such as modal synthesis before being incorporated into the multibody program. It is assumed by many analysts that the basis functions used are to a certain extent arbitrary. It is argued that so long as they are members of an infinite family of orthogonal functions and satisfy the geometric boundary conditions, convergence of the dynamic response is guaranteed.

It is well-known in the mathematical theory of variational methods that the basis functions must be complete in an appropriate inner product space. Our problem demonstrates the importance of this somewhat subtle criterion. It is extremely significant that application of the variational principle using an inappropriate set of basis functions is convergent, but to an erroneous result. This would be difficult if not impossible to identify in a typical spacecraft application

Problem Statement and Exact Solution

A cantilever beam constant bending stiffness EI and unit length is acted on by a uniform distributed loading. If the load per unit length is q, and we write p for the ratio q/EI, then the static deflection $W(x)$ is governed by

$$\frac{d^4 W}{dx^4} = p, \quad W(0) = W'(0) = W''(1) = W'''(1) = 0 \tag{1}$$

With constant p the solution is a polynomial of degree four:

$$W(x) = p\left(\frac{x^4}{24} - \frac{x^3}{6} + \frac{x^2}{4}\right) \tag{2}$$

For future reference we develop this solution in a Fourier cosine series, extending it as an even function to $(-1,0)$:

$$W(x) = p\left[\frac{1}{20} - \sum_{n=1}^{\infty} \frac{2 \cos n\pi x}{n^4 \pi^4} + \sum_{n=1}^{\infty} \frac{(-1)^n \cos n\pi x}{3n^2 \pi^2}\right] \tag{3}$$

Variational Solution

A variational principle equivalent to the boundary value problem (1) is the principle of minimum total potential energy: Find the function $W(x)$ satisfying $W(0) = W'(0) = 0$ (and sufficiently regular) that minimizes the total potential energy $EI\ P(W)$, where

$$P(W) = \int_0^1 \left[\frac{1}{2}(W'')^2 - pW\right] dx \tag{4}$$

We apply the Rayleigh-Ritz method to the energy functional in (4). The set of basis functions will be $\{1, \cos \pi x, \cos 2\pi x, \ldots\}$, which is complete over the interval $0 \le x \le 1$. The variational solution is then a cosine expansion

$$W = a_0 + \sum_{1}^{\infty} a_n \cos n\pi x \tag{5}$$

which automatically satisfies the geometric boundary condition $W'(0) = 0$ (as well as the natural boundary condition $W'''(1) = 0$). Substituting the assumed expansion (5) into the integral (4), the orthogonality of the cosines gives the value J for the total potential energy:

$$J = \frac{1}{4} \sum_{n=1}^{\infty} n^4 \pi^4 a_n^2 - p a_0 \tag{6}$$

We calculate the coefficients a_0, a_1, ... by rendering J stationary subject to the remaining geometric boundary condition: $W(0) - 0$, or $\Sigma a_n - 0$.

Following the standard method we introduce a Lagrange multiplier λ and build the constraint into the auxiliary function

$$L - \frac{1}{4} \sum_{n=1}^{\infty} n^4 \pi^4 a_n^2 - pa_0 + \lambda \left(a_0 + \sum_{n=1}^{\infty} a_n \right)$$

The equation $\partial L/\partial a_0 - 0$ yields $\lambda - p$. Then $\partial L/\partial a_n - 0$ gives

$$a_n - \frac{-2p}{n^4 \pi^4} \quad (n-1,2,3,\ldots) \tag{7}$$

The geometric constraint then yields

$$a_0 - \frac{2p}{\pi^4} \sum_{n-1}^{\infty} \frac{1}{n^4} - p/45 \tag{8}$$

When the coefficients (7), (8) are inserted into (5) we obtain the variational solution

$$W*(x) - p \left[\frac{1}{45} - \sum_{n=1}^{\infty} \frac{2 \cos n\pi x}{n^4 \pi^4} \right] \tag{9}$$

This can also be summed exactly, and it is again a fourth degree polynomial - but a different one!

$$W*(x) = p \left[\frac{x^4}{24} - \frac{x^3}{6} + \frac{x^2}{6} \right] \tag{10}$$

Comparing with (2), the difference $W-W*$ is $px^2/12$. Only geometric boundary conditions were enforced in computing $W*$. One expects that the natural boundary conditions will automatically be satisfied. Note however that $(W*)'' - - p/6$ at $x - 1$, and not zero. A plot of W (exact) and $W*$ (Ritz) is given in Figure 1.

BEAM DEFLECTION - RITZ SOLUTION
COSINE BASIS FUNCTIONS
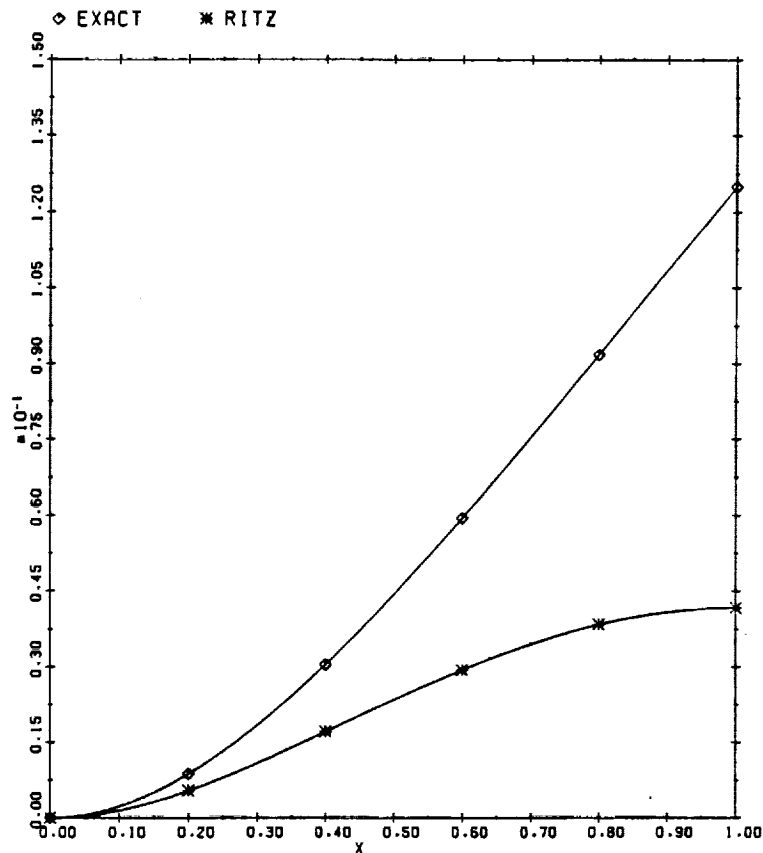
◇ EXACT   ✳ RITZ

Figure 1

An Observation

The paper started by expanding the exact solution W into the cosine series (3). That series has coefficients of order $1/n^2$, and it converges to W. But computing the bending energy in these two forms of the solution, polynomial and Fourier, leads to a disturbing result:

For the polynomial,

$$\int_0^1 (W'')^2 \, dx = p^2 \int_0^1 \left(\frac{1}{2} x^2 - x + \frac{1}{2} \right)^2 dx \quad \text{is } \textit{finite}.$$

For the series,

$$\int_0^1 (W'')^2 dx = p^2 \sum_1^\infty \frac{n^4 \pi^4}{2} \left[\frac{-2}{n^4 \pi^4} + \frac{(-1)^n}{3n^2 \pi^2} \right]^2 \quad \text{is } \textit{infinite}.$$

The orthogonality of the cosines produces this sum of positive terms, roughly $n^4 a_n^2$, and the sum does not converge.

834

Resolution of the Paradox

When the geometric boundary conditions are linear and homogeneous, it is standard practice to choose the basis functions such that each individually satisfies them. Since our basis functions of cosines do not all vanish at x=0, one might be tempted to point to this fact as the root of the problem. However, our enforcement of the geometric boundary condition through a Lagrange multiplier is perfectly legitimate. There is nothing in the theory that forces us to satisfy the geometric boundary conditions by each basis function; only the resultant linear combination must satisfy them. (See the section "Numerical Results" for substantiation of this statement.) The key to resolving the paradox lies in a loose statement preceding eq. (5), regarding the "completeness" of the cosines. The word "complete" in itself has no meaning. We have to identify the space of admissible functions, as well as its norm (the measure of distance in that space), before it can be claimed that a set of trial functions is complete – in other words, before we can say that the combinations of the trial functions can approximate with arbitrary accuracy any admissible function. To discuss accuracy we need a norm.

The most common measure of distance is the L^2 norm – the square root of $\int f^2 dx$. The function f need not be continuous; step functions present absolutely no difficulty. The space contains functions much worse than that, although a delta function has infinite length and is not allowed. In the L^2 norm the set $\{1, \cos \pi x, \cos 2\pi x, \ldots\}$ is *complete* (on the interval $0 \leq x \leq 1$) and this is a cornerstone of Fourier analysis. The cosines are also complete in the L^p spaces, with norm $(\int |f|^p dx)^{1/p}$ and $p \geq 1$, but L^2 is special: it is associated with an inner product. That makes it a Hilbert space, in which $(f, g) = \int fg \, dx$ matches the norm: (f,f) agrees with $\|f\|^2$. The L^2 space admits angles, and orthogonality, and all the geometry of ordinary Euclidean space. But it is not the only Hilbert space, nor is it necessarily the right one.

Our fourth-order problem comes with its own norm and inner product and space of admissible functions. We look there for the resolution of the paradox; we have to work with the right space. The *norm* comes directly from the bending energy:

$$\|W\|^2 = \int_0^1 (W''(x))^2 dx$$

The inner product is determined by the norm:

$$(W,V) = \int_0^1 W''(x)V''(x)dx$$

The admissible functions are also determined: Their norm must be finite and they must satisfy the essential boundary conditions. Thus W is admissible if

$$\int_0^1 (W'')^2 dx < \infty \quad \text{and} \quad W(0)=0 \quad \text{and} \quad W'(0)=0$$

835

Such a function W comes from integrating twice a function in L^2:

$$W(x) = \int_0^x \int_0^y f(t)dt\, dy$$

Notice that W is not required to satisfy the natural conditions $W''(1)=0$ and $W'''(1)=0$. We could not make that requirement and still have a complete space. Functions that satisfy these extra conditions can come arbitrarily close (measured by the norm) to functions that don't. The process of completion wipes out the natural conditions as a requirement on admissible W. In Reference 1 the second author referred to the admissible space as H_E^2 – the Hilbert space of functions that have two derivatives (in $L^2 = H^0$) and that satisfy the essential boundary conditions. Remark 3 will justify more fully the choice of bending energy – the second-degree term in the total potential energy – as the norm.

That finishes the functional analysis. It was needed in order to ask the right question: *Are the cosines complete in the space of admissible functions?* We suspect that the answer must be no.

Apart from boundary conditions, we are asking whether combinations of the cosines (including cos 0=1) can come arbitrarily close to W. We know they can do so in the ordinary L^2 norm, and the Fourier expansion (3) does it explicitly. The question is whether the cosines can come arbitrarily close in the 'second-derivative norm'. Equivalently, the second derivatives of our set of cosines must come close, in the ordinary sense to W''. But the second derivatives are

$$\{0,\ -\pi^2 \cos \pi x,\ -4\pi^2 \cos 2\pi x, \ldots\}$$

You see the problem. We are missing the constant term! We cannot approximate $W''=2$ with the functions we have left. In other words, we cannot come close to $W=x^2$ with our original set of trial functions. The cosines were not complete, but if this additional trial function x^2 is included, the set is complete. That would add the constant function to the list of second derivatives. So it was no accident that the discrepancy between W and W* was a multiple of x^2.

Before drawing a final conclusion about W* we add four observations.

1. The new function x^2 not only completes the set, it is orthogonal to the original cosines. The inner product is

$$\int_0^1 -n^2\pi^2 \cos n\pi x \cdot 2\, dx = 0$$

2. The original cosines were not in H_E^2 – they were not really admissible – because they violated the essential condition $W(0)=0$. One way to correct that would have been to construct the trial functions more carefully; they could have been

$$\{1 - \cos \pi x, \ 1 - \cos 2\pi x, \ 1 - \cos 3\pi x, \ldots\} \quad \text{(and also } x^2!)$$

The alternative of keeping the extra trial function 1, and imposing $W(0)=0$ through a Lagrange multiplier, is equally correct. The example verifies that the choice can be based on computational convenience. In other problems it might be less easy to adjust the trial functions to satisfy the essential conditions.

3. The special feature of the norm $\|W\|^2 = \int_0^1 (W'')^2 \, dx$ is that the distance from any admissible w to the exact solution W satisfies

$$\frac{1}{2} \| w - W \|^2 = P(w) - P(W) \tag{11}$$

It follows that minimizing the potential energy P over all trial functions w automatically minimizes the distance to W. If the trial space actually contains W, the distance is zero and we have the global minimum of P. In the typical case, when the Rayleigh-Ritz method keeps w in a finite-dimensional space, the best w is the projection of the exact W onto the trial space. The w that minimizes P also has minimum error – but to establish (11) we must measure the error $\|w - W\|$ in the correct 'energy norm'.

The verification splits off a term of integration by parts:

$$\| w - W \|^2 = \int_0^1 [(w'')^2 + 2W''(W'' - w'') - (W'')^2] \, dx$$

$$\int_0^1 W''(W'' - w'') \, dx = \int_0^1 \frac{d^4 W}{dx^4} (W - w) \, dx + [W''(W' - w') - W'''(W - w)]_0^1$$

the boundary terms vanish because of the essential conditions on W and w at $x=0$, and the natural conditions on W at $x=1$. Writing p for d^4W/dx^4, and substituting into the first line, we recognize its right hand side as $2P(w) - 2P(W)$. The energy norm imposes itself. In the 'energy inner product', there is an orthogonal projection of W onto the trial space. That is why Rayleigh-Ritz, and finite elements, do so well.

4. A final small worry. If a set is not complete, as the original cosines were not, the terms they give should satisfy Bessel's inequality:

$$\|\Sigma a_n \cos n\pi x\|^2 \leq \|W\|^2$$

This means that the component of W inside the trial space should not be larger than W itself. But the inequality was tested in the short section prior to this one, and it failed. The left side was $+\infty$ and the integral of $(W'')^2$ on the right side was finite. After some thought one realizes that the a_n may be the cosine coefficients of W, but they are not the coefficients in the right inner product! Bessel's inequality with the series (3) is satisfied in the L^2 norm. In the energy norm, we must use the coefficients (7), (8) obtained from the Rayleigh-Ritz method.

Summary

It follows from the above discussion that the Rayleigh-Ritz method has converged upon the best approximation (as measured in the energy norm) in the space spanned by the cosines. We can write the approximation (9) in the alternate form

$$W* = \sum_{n=1}^{\infty} \frac{2p}{n^4 \pi^4} (1 - \cos n\pi x)$$

The above is not the cosine expansion of W (eq. 3). The two are different because the x^2 term was forgotten. Without that term we have Bessel's inequality $\|W*\|^2 \le \|W\|^2$. With x^2 included to complete the set of trial functions, the variational solution will become the exact one:

$$W = \frac{p}{12} x^2 + \sum_{1}^{\infty} \frac{2p}{n^4 \pi^4} (1 - \cos n\pi x)$$

We can verify that $p/12$ is the Fourier coefficient in the correct inner product:

$$\frac{(W, x^2)}{(x^2, x^2)} = \frac{\int W'' \cdot 2dx}{\int 2 \cdot 2dx} = \frac{p}{2} \int_0^1 \left(\frac{x^2}{2} - x + \frac{1}{2} \right) dx = \frac{p}{12}$$

This example serves to illustrate an elusive pitfall in practice. It is possible to see rapid numerical convergence and conclude that a good approximation to the exact solution has been obtained. In reality, if one's set of basis functions are not complete (in the appropriate inner product space) the approximate solution may be highly inaccurate. Finite element experts will note that the pitfall could have been avoided by applying the *patch test*. That requires the special solution $W=x^2$, with constant strain, to be reproduced by the trial functions — after imposing conditions on an element boundary consistent with this particular W. The cosines would not have reproduced x^2. Thus the patch test would have failed, correctly indicating that the set was incomplete.

The reader will have noticed from the beginning that all combinations of the cosines satisfy the extra condition $W'(1)=0$. If that had been the essential condition at the right-hand end (see Reference 2, p. 174), with $W'''(1)=0$ as natural condition, the cosines would have been adequate.

Note: A completely parallel example can be constructed for the second-order equation -u"=1, with essential condition u(0)=0 and natural condition u'(1)=0. The corresponding set of trial functions, complete in L^2 but incomplete in H_E^1, is {sin nπx}.

Numerical Results

Here we compare the cosine expansion to a different expansion that the theory guarantees to succeed. The latter comes from the eigenfunctions of a uniform vibrating free-free beam:

$$\frac{d^4\phi}{dx^4} = \beta^4\phi \quad \text{with} \quad \phi''(0) = \phi'''(0) = \phi''(1) = \phi'''(1) = 0 \tag{12}$$

Note that a more natural choice is a clamped-free beam, satisfying $\phi(0)=\phi'(0)=0$ at the left endpoint (and more like 1 - cos nπx). We wanted to see how the Lagrange multipliers would enforce these geometric conditions, when they are not imposed on each eigenfunction.

The results are striking. For the cosine expansion (through cos Nπx) we tabulate the deflections at x=1 and their errors (see Table I). The exact value is W*(1)=p/24. W* will be supplemented by the $px^2/12$ correction term, which is twice as large! Together they reach the correct value W(1)=p/8.

What is significant is the $1/N^3$ convergence rate. Doubling N reduces the error in the last column by a factor of 8.

Contrast that with the results using the free-free eigenfunctions

$$\phi_1 = 1 \quad \text{with} \quad \beta_1 = 0$$

$$\phi_2 = x - \frac{1}{2} \quad \text{with} \quad \beta_2 = 0$$

$$\phi_n = (\sin \beta_n - \sinh \beta_n)(\cos \beta_n x + \cosh \beta_n x)$$

$$- (\cos \beta_n - \cosh \beta_n)(\sin \beta_n x + \sinh \beta_n x)$$

$$\text{with} \cos \beta_n \cosh \beta_n = 1$$

The frequencies β_n are asymptotic to πn + constant. The expansions can almost be carried out by hand (with the help of orthogonality), but imposing W(0)=W'(0)=0 by Lagrange multipliers needs a simple code. The deflections are again tabulated at the free end x=1, and you will notice the change in convergence rate (see Table II). The error decreases like 1/5N. At N=800 we are far above the error achieved previously at N=10.

	Table I.	
N	$W*_N(1)/p$	Error
10	0.04165996	0.00000671
20	0.04166582	0.00000085
40	0.04166656	0.00000011
80	0.04166665	0.00000002
Exact	0.04166667	

	Table II.	
N	$W_N(1)/p$	Error
50	0.120908	0.004092
100	0.122966	0.002034
200	0.123988	0.001012
400	0.124496	0.000504
800	0.124750	0.000250
Exact	0.125000	

Finite Elements

This last section looks at the effect of enforcing natural boundary conditions (as well as essential conditions) in the finite element method. The trial functions will be piecewise polynomials, and for the bending problem the natural choice is piecewise cubics. There are two parameters at each node – displacement and slope. Because they are continuous between elements, the trial functions w are in the class C^1 – with one continuous derivative, and jump discontinuities in the second derivative. This guarantees that w will be conforming; its bending energy (and therefore its norm!) is finite. We have only to think about the boundary conditions.

The essential conditions fix $w_0=0$ and $w'_0=0$ at the left end. With N intervals of length $h=1/N$, and two degrees of freedom per node, that leaves a 2N-dimensional trial space – provided no conditions are applied at the right end. If we do impose the natural boundary conditions, they yield two relations between w_{N-1}, w'_{N-1}, w_N, w'_N. Therefore the trial space is reduced to dimension 2N -2. The question is whether this is wise – to compel w to meet conditions that we know to be satisfied by the true solution W.

A functional analyst would say it is foolish. The smaller trial space (it is a subspace of the larger one) cannot give a better approximation to W. The error must be greater when degrees of freedom are removed. But that is the error in the energy norm, where the identity (11) holds and Rayleigh-Ritz picks the best finite element approximation as the projection. We might still hope that pointwise, and particularly near $x=1$, there is something to be gained by imposing the natural conditions.

This example is so simple that the calculations can be done with pencil and paper. Furthermore, it has the special property of *superconvergence*. The finite element approximation is *exact at the nodes*. In the full trial space, with no constraints at $x=1$, there is agreement with the true W at every meshpoint. In the smaller trial space, the natural boundary conditions require the cubic to be a linear function within the final interval. (Then it satisfies $w''=w'''=0$ throughout the interval; that is the price for imposing those conditions at $x=1$.) In this case it is still exact at all other nodes! The discrepancy between the two finite element solutions is small, and very local, but the winner is clear.

Even near x=1, it is better when the natural conditions are left alone. By satisfying them, we spoil the accuracy.

It is a pleasure to verify superconvergence in this case. The element stiffness matrix and element force vector are

$$
\frac{1}{h^3}
\begin{bmatrix}
12 & 6h & -12 & 6h \\
6h & 4h^2 & -6h & 2h^2 \\
-12 & -6h & 12 & -6h \\
6h & 2h^2 & -6h & 4h^2
\end{bmatrix}
\quad \text{and} \quad
\frac{ph}{12}
\begin{bmatrix}
6 \\
h \\
6 \\
-h
\end{bmatrix}
$$

Those refer to the local parameters w_{j-1}, w'_{j-1}, w_j, w'_j. The assembly combines the overlapping parts:

$$
\begin{bmatrix} \begin{bmatrix} \quad \end{bmatrix} \end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix} \begin{bmatrix} \quad \end{bmatrix} \end{bmatrix}
$$

Therefore the global stiffness equations for $j<N$ (found from $\partial P/\partial w_j = \partial P/\partial w'_j = 0$) come from the middle rows of that assembly:

$$
-12w_{j-1} - 6hw'_{j-1} + 24w_j - 12w_{j+1} + 6hw'_{j+1} = ph^4
$$

$$
6hw_{j-1} + 2h^2 w'_{j-1} + 8h^2 w'_j - 6hw_{j+1} + 2h^2 w'_{j+1} = 0
$$

(13)

Those equations are exactly satisfied by the nodal values W_j and W'_j of the true solution W in equation (2).

At the end $x=1$, imposing $w''_N = 0$ and $w'''_N = 0$ leads to

$$
w'_N = w'_{N-1} \quad \text{and} \quad w_N = w_{N-1} + hw'_{N-1}
$$

(14)

The cubic w is forced to be linear in the last interval. On the other hand, if these natural conditions are not imposed, the only difference from (13) is that no terms are assembled for the interval beyond the boundary:

$$
-12w_{N-1} - 6hw'_{N-1} + 12w_N - 6hw'_N = ph^4/2
$$

$$
6hw_{N-1} + 2h^2 w'_{N-1} - 6hw_N + 4h^2 w'_N = -ph^5/12
$$

(15)

The latter are satisfied by the true W but (14) is not — even though W itself does satisfy the natural conditions.

Remark. Even if the right side of $W'''' = p$ is not constant in the original differential equation, it is still true that the finite element approximation has $w=W$ and $w'=W'$ at the nodes. This was observed already by Pin Tong[3]. Those

841

conditions determine one particular interpolating w – and to show that this special w is the finite element solution, we verify that its error w – W is orthogonal to all trial functions. Then w is the correct projection of W, and superconvergence is verified.

Proof of orthogonality: In each interval the trial functions are cubics C(x), and the contribution to the inner product is an integral over that interval:

$$\int(w-W)"C"dx = [(w - W)'C" - (w - W)C'''] $$

$$+ \int(w - W)C""dx \tag{16}$$

this is zero because w-W and (w-W)' are zero at the endpoints, and C"" vanishes identically.

The argument still applies when we impose the natural boundary conditions, and force w to be linear in the last interval. Over all other intervals (16) is zero as before. Over the last interval it is zero because C"=0. Therefore the cubic which matches W and W' at every node except x=1 is the finite element solution in this case. It would have been more accurate to match at *every* node, by not forcing w"(1)=w'''(1)=0.

In this example, and surely in more realistic and more complicated applications of the displacement method, *it is better to leave the natural boundary conditions to nature*.

References

1. G. Strang and G. Fix, <u>An Analysis of the Finite Element Method</u>, Prentice-Hall, Englewood Cliffs, NJ, 1973.

2. G. Strang, <u>Introduction to Applied Mathematics</u>, Wellesley-Cambridge Press, Box 175, Wellesley, MA, 1986.

3. Pin Tong, "Exact Solution of Certain Problems by the Finite Element Method", AIAA J. 7, 178-180 (1969).